



Universidad del Valle
Programa de Economía

Proyecto de Grado
Análisis de eficiencia de la educación media en Colombia.
Descomposición de efectos a través de técnicas no
paramétricas.

Estudiante
Gloria Rosaura Galvez Galvez
gloriarosagalvez@hotmail.com

Director
Alexei Arbona Estrada
aarbona@javerianacali.edu.co

Mayo, 2018

Resumen

Este trabajo mide la eficiencia (ineficiencia) de las escuelas públicas en Colombia estimando cuánto de esta eficiencia es atribuible a los estudiantes y cuánto a las escuelas. Metodológicamente se utiliza un enfoque de metafrontera (eficiencia) y se hacen estimaciones con order-m, una técnica no paramétrica con un componente aleatorio (bootstrapping). Se utiliza una función de producción educativa estándar con orientación al output que relaciona el desempeño de los estudiantes de educación media en escuelas públicas en Lectura y Matemáticas de acuerdo con la prueba Saber 11, con el nivel socioeconómico de los estudiantes y de la escuela, y la calidad académica de la misma. Se utiliza una base de datos de 139.045 estudiantes y 1.320 escuelas en los cuatro principales departamentos de Colombia que, en el año 2016, representan el 46.6% de los estudiantes de último año escolar en colegios oficiales. Se encuentra que el efecto estudiante domina al efecto escuela y que Barranquilla y Atlántico son las regiones donde, a pesar de tener los más bajos desempeños en las pruebas, sus estudiantes y, sobre todo, sus escuelas tienen los menores coeficientes de ineficiencia y, por tanto, utilizan mejor sus dotaciones de inputs. Este trabajo aporta a una mejor asignación de los recursos dirigidos a educación.

Palabras clave: Educación, Eficiencia, order-m

Clasificación JEL: I21, D61

Contenido

Resumen.....	i
Lista de Tablas.....	iii
Lista de Figuras.....	iv
1. Introducción.....	1
2. Literatura.....	3
3. Metodología.....	5
4. Datos.....	11
5. Resultados y discusión.....	16
6. Conclusiones y Recomendaciones.....	23
Referencias.....	26

Lista de Tablas

Tabla 1. Descripción de la muestra (escuelas y estudiantes por subregión). Colombia, 2016	13
Tabla 2. Descriptivos: <i>inputs</i> y <i>outputs</i> por subregiones (Colombia, 2016)	14
Tabla 3. Coeficientes de eficiencia por efectos: OE=STExSCE	18
Tabla 4. Resumen descomposición de efectos OE=STExSCE	21

Lista de Figuras

Figura 1. Descomposición de la ineficiencia de los estudiantes.....	7
Figura 2. Descomposición empírica de la ineficiencia de los estudiantes.....	7
Figura 3. Índice de condiciones socioeconómicas y culturales (Análisis de Correspondencia Múltiple)	11
Figura 4. Desempeño municipal en lectura y matemáticas (promedio por área, Saber 11, ICFES, Colombia 2016).	15
Figura 5. Descomposición de efectos OE=STExSCE (Colombia 2016).....	20
Figura 6. Peso relativo de los efectos estudiante y escuela (Colombia, 2016).....	22
Figura 7. Casos atípicos en la descomposición: SCE por municipios (Valle y Atlántico, 2016).....	23

1. Introducción

El objetivo de este trabajo es estimar la eficiencia de las escuelas públicas en Colombia. Responder qué tan eficiente es el sistema educativo colombiano contribuye a una mejor comprensión de los determinantes del desempeño educativo y a una mejor asignación de los recursos. Este trabajo, además, descompone los niveles de eficiencia entre aquel que se le puede atribuir a los estudiantes y el atribuible a las escuelas. Lo anterior es importante tanto para el diseño de la política pública sectorial, como para la gestión educativa de las escuelas públicas e incluso para orientar las inversiones que, voluntariamente, realizan otros actores de la sociedad civil.

En la perspectiva de eficiencia, esto significa qué tanto aprovechan sus dotaciones los estudiantes y las escuelas públicas frente a su desempeño académico. Y, por tanto, qué potencial de mejora tiene cada unidad de análisis frente a grupos de referencias comparables (alumnos, escuelas o territorios). En el desarrollo de esta pregunta de investigación, este trabajo también responde si estas medidas de eficiencia presentan diferencias inter e intra regionales para las cuatro regiones evaluadas: Bogotá y Cundinamarca, Antioquia, Valle del Cauca, y Atlántico y sus municipios frente a sus ciudades capitales, donde es importante resaltar que este trabajo distingue en los departamentos a sus respectivas ciudades capitales para controlar las variables de entorno como el tamaño de las ciudades o la capacidad institucional, entre otros. Este trabajo asume como hipótesis, de conformidad con la literatura, que los factores asociados a los estudiantes afectan en mayor medida la eficiencia total frente a los que están asociados a las escuelas y que, en consecuencia, una política

óptima asignaría mayores recursos a esos factores más próximos al entorno individual y familiar del estudiante en contra de lo que es observado.

Metodológicamente, la estimación de la ineficiencia total se realiza a través de una técnica de frontera no paramétrica denominada *order-m* la cual es un refinamiento del análisis envolvente de datos (DEA), aplicando un componente aleatorio de remuestreo o de *bootstrapping* para controlar valores extremos y atípicos, algo muy frecuente en resultados académicos. En segundo lugar, para distinguir los efectos estudiante y escuela se realiza una descomposición usando el enfoque de metafronteras. La combinación de ambas técnicas es relativamente reciente y novedosa en la literatura de eficiencia.

Este estudio concluye con tres grandes resultados: i) Existe un nivel de ineficiencia en los colegios públicos entre 20%-35% para las 8 submuestras del estudio (4 departamentos y sus respectivas ciudades capitales) ii) Salvo Bogotá, la ineficiencia del resto (departamento sin su respectiva ciudad capital) es mayor a la ineficiencia de las capitales; en otras palabras, se encuentra que las regiones tienen un desempeño diferente entre ellas, pero además una diferencia entre ellas y sus ciudades capitales y iii) el efecto estudiante es mayor que el efecto escuela y toma valores entre 57% y 89% lo cual es consistente con la literatura.

Por último, este documento, después de esta sección introductoria, incluye una revisión de literatura (sección 2), define el modelo de análisis (sección 3), una descripción de los datos utilizados (sección 4), los resultados obtenidos y su discusión (sección 5) y una sección final de conclusiones y recomendaciones (sección 6).

2. Literatura

Los retornos positivos de la educación sobre las personas (Psacharopoulos, 1981) y sobre los niveles de desarrollo de los países (Hanushek & Woessmann, 2008; Psacharopoulos & Patrinos, 2004) la siguen convirtiendo en una de las principales estrategias de intervención para los hacedores de política y en tema de investigación para los académicos. Mayores ingresos en los mercados de trabajo (Hanushek & Kimko, 2000), mayores niveles de crecimiento económico (Barro, 2001; Hanushek & Wößmann, 2010) y mejoras en la equidad (Coleman, 1966; Gregorio & Lee, 2002) son algunos de los efectos más estudiados.

El gasto de los gobiernos y de las familias responde a esa evidencia de la literatura. Según la OECD (2017), los gobiernos gastan en educación, en promedio, 5.5% de su PIB mientras los hogares gastan 2.5% adicionales. Estudiar la eficiencia del sistema educativo (las escuelas y los estudiantes, en este caso) brinda elementos para mejorar la eficiencia asignativa del gasto y estimular los impactos positivos de una mejor educación.

A partir del famoso y extenso Reporte Coleman (1966) se fija en la literatura una orientación a los resultados (outputs y outcomes) de la educación que sentó las bases para la formulación de las primeras funciones de producción educativa (Bowles, 1970; Hanushek, 1979, 1986, 2008), en torno a las cuales aun en la actualidad hay un gran consenso, aun a pesar de que desde ese trabajo se reconoció un peso y mayor y más influyente al entorno familiar que al papel de la escuela. No obstante, el papel de la escuela sigue siendo relevante y existe una corriente basada en ella (*school-based literature*) que sigue generando un alto volumen de publicaciones. Tomando como referencia ERIC (Education Resources Information Center, la mayor base de datos especializada en educación, financiada por el Institute of Education Sciences of U.S. Department of Education), la búsqueda ‘school-based’ lista 396

artículos publicados en revistas científicas arbitradas durante los últimos 5 años en la categoría ‘Program Effectiveness’ y 269 en ‘Academic Achievement’ (<https://eric.ed.gov/>, fecha de consulta: mayo 3 de 2018).

La escuela sigue siendo el lugar en el que sucede la mayor parte de los procesos educativos y la que atrae la mayor cantidad de recursos asignados a la calidad educativa (Hanushek, 2006), luego, la pregunta sobre eficiencia educativa y la contribución de la escuela a ella sigue siendo relevante (Angrist, Pathak, & Walters, 2013). El impacto de mejores escuelas (con mayores recursos y con un mejor aprovechamiento de ellos) es positivo frente al desarrollo (Hanushek & Woessmann, 2008). No obstante, también hay suficiente evidencia del aporte del entorno familiar de los estudiantes a su mejor desempeño académico (Martins & Veiga, 2010; Muhammad, Iqbal, & Tasneem, 2015).

Metodológicamente, entre las distintas formas de abordar la tarea de evaluar con evidencia el mejor desempeño del alumno y de la escuela, se destacan dos líneas en la literatura empírica: (i) las evaluaciones de impacto usando métodos cuasi experimentales y experimentos aleatorizados y (ii) las técnicas de frontera no paramétricas usando un enfoque de eficiencia.

Este trabajo elige un enfoque de eficiencia por varias razones. En primer lugar, porque permite utilizar múltiples inputs para producir múltiples outputs: *“The presence of multiple inputs, multiple outputs, and prices that are unlikely to serve a useful purpose as weights, are features that combine to make DEA an instructive tool in this context”* (Thanassoulis et al., 2016, p. 368). Adicionalmente, este enfoque no impone supuestos sobre la forma funcional ni sobre la distribución de los errores. En una función de producción como la educativa donde hay tantos elementos no observables (vg. habilidad, compromiso, motivación de estudiantes, profesores o padres de familia), esto es relevante. Finalmente, la

frontera empírica con base en los datos observados facilita la comparación de los estudiantes con aquellos que son más parecidos a él. De Witte & López-Torres (2015) brindan una extensa revisión de literatura en eficiencia en educación.

Sobre la técnica, order-m, que es una técnica no paramétrica (Deprins, Simar, & Tulkens, 1984) con el enfoque de Análisis Envolvente de Datos (DEA) aplicando un refinamiento con un componente aleatorio de remuestreo (Cazals, Florens, & Simar, 2002; Simar, 2003) para dar manejo a los valores extremos (una crítica habitual a las técnicas de eficiencia). Para la descomposición de los efectos se hace una aplicación del enfoque metafronteras, como lo sugieren Silva Portela & Thanassoulis (2001) y Thieme, Prior, & Tortosa-Ausina (2013). Una aplicación reciente de este enfoque para el caso colombiano, utilizando datos de PISA 2012, puede ser consultado en de Jorge-Moreno, Díaz, Rodríguez, y Segura (2018). En este trabajo, como en el trabajo previo de Thieme et al. (2013), la evidencia es la misma: el efecto estudiante domina al efecto escuela.

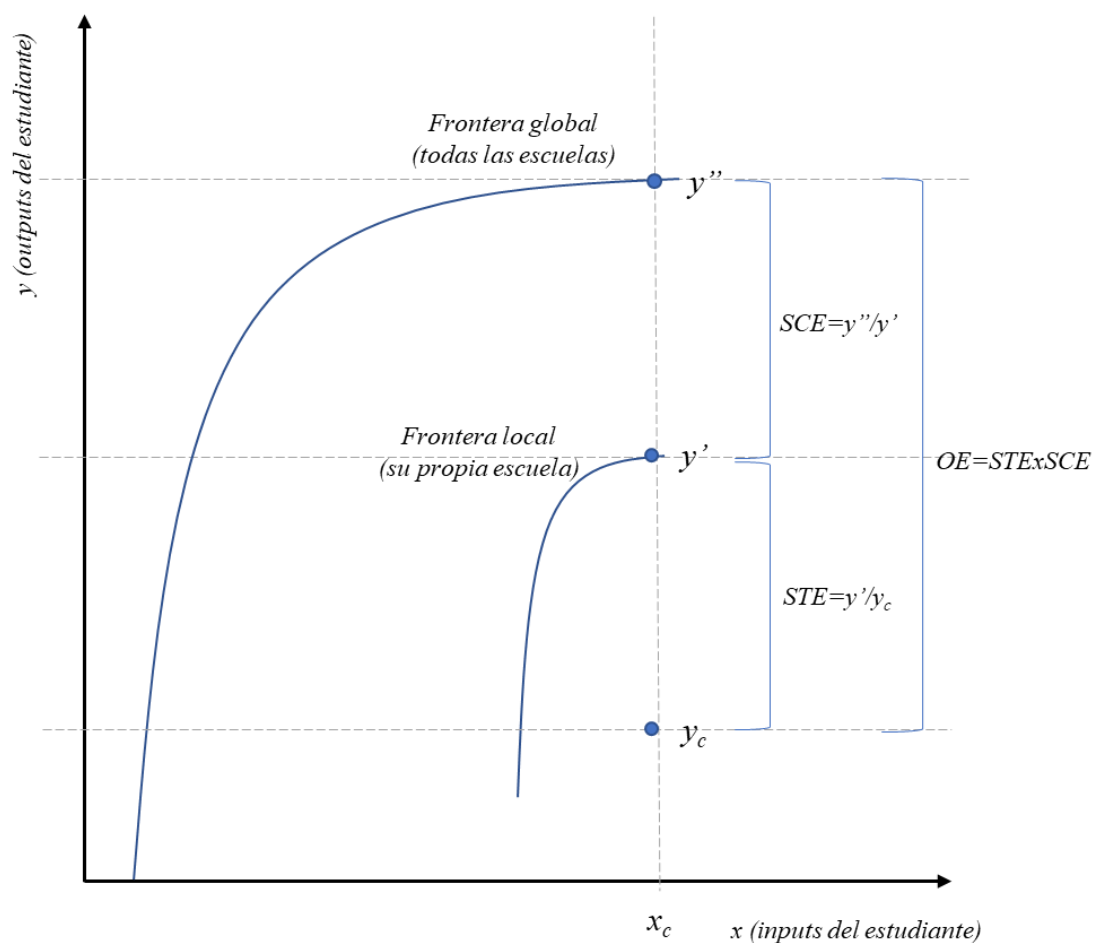
3. Metodología

Esta metodología propone la estimación de al menos un par de fronteras para estimar la distancia entre ellas como medida del efecto marginal de una con respecto a la otra adaptando el enfoque propuesto por Thanassoulis & Silva Portela (2002) y desarrollado Thieme et al. (2013).

En este caso, como se ilustra en la Figura 1 y dada la orientación al output, se trata de evaluar la eficiencia (ineficiencia) de un estudiante c que, dado un nivel de input x_c , podría maximizar su nivel de output y_c , frente a su frontera de referencia. En el primer caso, el estudiante se compara contra una frontera global y'' (los mejores estudiantes de todas las

escuelas del grupo de referencia con un nivel de input comparable). En el segundo caso, el estudiante se compara contra una frontera local y' (los mejores estudiantes de su escuela con un nivel de input comparable al suyo). Las distancias y''/y_c y y'/y_c representan entonces una medida de la ineficiencia del estudiante c frente al respectivo grupo de referencia. A la distancia y''/y_c que mide la ineficiencia total se le denomina *Efecto Total* ($OE = y''/y_c$) y en ella se incluye la totalidad de las fuentes de ineficiencia o el conjunto amplio de determinantes (observables y no observables). En tanto, la distancia y'/y_c que mide la ineficiencia local se denomina *Efecto Estudiante* ($STE = y'/y_c$) y a ella se le atribuyen las fuentes de ineficiencia que están al interior del colegio. Se le denomina efecto estudiante porque, para ese grupo, todos los alumnos reciben los mismos inputs por parte del colegio y sólo cambian los propios y los de su grupo familiar. Recibiendo los mismos inputs, el potencial de mejora es plenamente atribuible a él. Al efecto residual, en consecuencia, se le atribuyen las fuentes de ineficiencia asociadas a las diferencias entre escuelas. Este trabajo denomina a este efecto, *Efecto Estudiante* (STE). Por construcción, dado que $OE = STE \times SCE$, entonces $SCE = OE/STE$ de tal forma que reemplazando los términos $STE = y''/y'$.

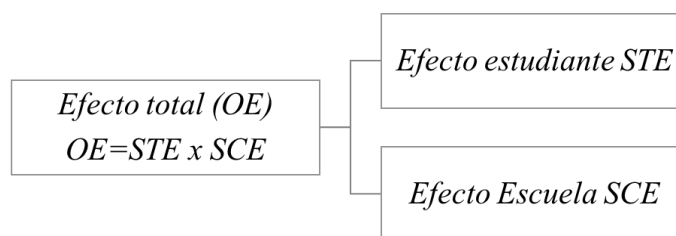
Figura 1. Descomposición de la ineficiencia de los estudiantes



Fuente: (Thanassoulis & Silva Portela, 2002 en Thieme et al., 2013)

En resumen, tal como se muestra en la Figura 2, el OE se descompone en SCE y STE

Figura 2. Descomposición empírica de la ineficiencia de los estudiantes



Empíricamente, para estimar los coeficientes de los efectos se utiliza *order-m*, una técnica no paramétrica de frontera con un enfoque de eficiencia (Cazals et al., 2002; Daraio & Simar, 2005). El enfoque de estimación *order-m*, comparte con su antecesor, el Análisis Envolvente de Datos (DEA), la capacidad de evaluar la eficiencia relativa de unidades homogéneas de toma de decisiones (DMUs) frente a las fronteras de eficiencia, utilizando más de un output sin tener que asumir una forma funcional o establecer supuestos sobre la distribución de los errores (Charnes, Cooper, & Rhodes, 1978). A diferencia de los modelos DEA, pero coincidiendo con la variante Free Disposal Hull (FDH) (Deprins et al., 1984), *order-m* no asume convexidad, es decir, que sólo son posibles comparaciones con DMUs observables y no con combinaciones lineales de ellas. *Order-m* consigue además ser menos sensible a valores extremos y atípicos y no sufre el problema de *maldición de dimensionalidad*.

Complementariamente a otras técnicas de frontera eficiente, en el proceso de optimización de dicha medida de eficiencia *order-m* define una muestra aleatoria con reemplazo de tamaño m en la que compara el nivel de output observado en cada DMU ($y_{c,j}$) con m DMUs aleatorias que cumplan la condición de tener un nivel de output mayor que el observado [$y_{m,j} > y_{c,j}$ de m random variables ($y_{m,j}, \dots, y_{m,j}$)]. A esta medida de eficiencia de dicha muestra de referencia artificial se le conoce como pseudo FDH-eficiencia $\theta_{mi}^{FDH_b}$. A continuación, la técnica introduce un componente de bootstrapping realizando B veces este procedimiento de tal forma que la medida de eficiencia resulta de la medida de las B submuestras (Daraio & Simar, 2007).

$$\theta_{mi}^{\widehat{OM}} = \frac{1}{B} \sum_{b=1}^B \widehat{\theta}_{mi}^{FDH_b}$$

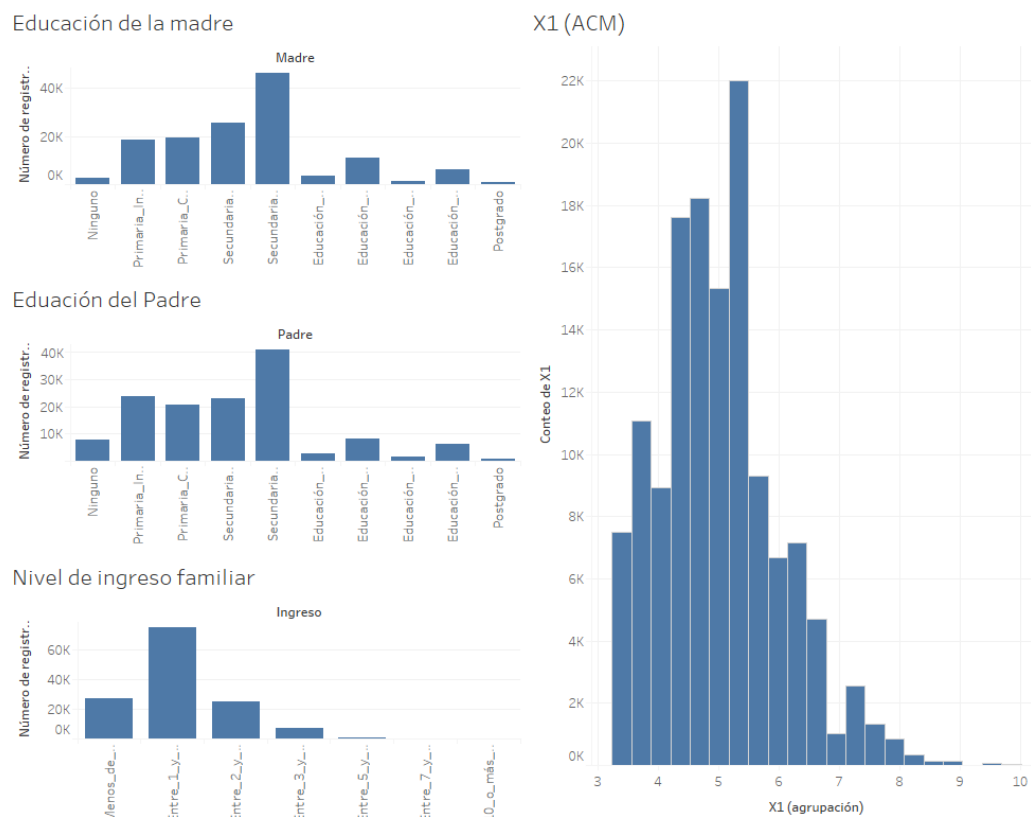
Debido a que FDH es sensible a los valores atípicos, entre más grande sea m , es más probable que dichas observaciones se incluyan en la muestra y más se aproximará *order-m* a la estimación FDH. Por el gran tamaño de cada muestra y las diferencias de tamaño entre ellas, se definió m igual al 1% de la muestra en la estimación de fronteras globales (todos los estudiantes de una región). Para las fronteras locales (los colegios), se definió un $m=30$; este valor coincide con la restricción que este trabajo impuso para la conformación de la muestra en la que sólo se evalúa la eficiencia de estudiantes que están en colegios con al menos 30 estudiantes. El parámetro B es igual a 200 (pruebas con valores superiores no mostraron cambios de importancia en la magnitud de los coeficientes, pero sí en la duración de los procesos computacionales). Esta condición de remuestreo para la estimación de los coeficientes de eficiencia y los tamaños de m y B imponen unas exigencias computacionales reportadas en la literatura (Tauchmann, 2012) y que este trabajo ha resuelto utilizando ‘*nonparaeff*’ package para R, algunos códigos de la rutina de ‘doparallel’ para optimizar el uso de recursos (procesadores) en paralelo. Esta es una de las razones por la cual no es usual encontrar en la literatura estimaciones de esta naturaleza con tamaños de muestra equivalentes: 47,076 observaciones en Thieme et al. (Thieme, Prior, Tortosa-Ausina, & Gempp, 2016), 22,313 Cordero, Prior, et al. (Cordero, Prior, & Simancas, 2015) y 11,319 en Thieme et al. (Thieme et al., 2013), entre otros.

La función de producción elegida es una convencional en la literatura sobre eficiencia en educación. Una revisión exhaustiva de, entre otros, las variables más utilizadas ha sido realizada por (De Witte & López-Torres, 2015).

Este trabajo, los estudiantes de último año escolar son los DMUs, y se definen como outputs sus puntajes en Lectura (y_1) y Matemática (y_2) en una prueba estandarizada. Los inputs definen tres fuentes de eficiencia (ineficiencia): al nivel socioeconómico y cultural del

estudiante (x_1), la calidad del colegio al que asiste (x_2) y el nivel socioeconómico y cultural de sus compañeros (x_3). La primera variable es a nivel del estudiante y las otras dos a nivel del colegio. Para definir x_1 se estima una variable latente a través de un Análisis de Correspondencia Múltiple (ACM) (Tenenhaus & Young, 1985) considerando los años de educación de la madre, los años de educación del padre y el ingreso familiar mensual (Ver Figura 3). La media de calidad del colegio x_2 emplea el índice Sintético de Calidad (ISCE) correspondiente al nivel medio de educación (décimo y onceavo grado escolar). Este índice es calculado por el ICFES (la autoridad nacional de evaluación educativa) y para el nivel de educación media pondera tres componentes: progreso, desempeño y eficiencia. Por construcción, x_2 considera en uno de sus componentes el desempeño de cada colegio en los años previos al año evaluado a partir del desempeño de otras cohortes de estudiantes. Finalmente, x_3 se calcula promediando x_1 para cada colegio. A diferencia de trabajos previos en que cada input se añade progresivamente para ver su contribución (Thieme et al., 2013). Este trabajo mantiene constante la especificación del modelo y_1 , y_2 (*outputs*), x_1 , x_2 , x_3 (*inputs*) y, en cambio, evalúa submuestras de referencia diferentes para estimar las fronteras con las que se calculan los efectos.

Figura 3. Índice de condiciones socioeconómicas y culturales (Análisis de Correspondencia Múltiple)



Fuente: Cálculos propios

4. Datos

Tres fuentes fueron utilizadas para la consolidación de la base de datos con la que se hacen las estimaciones en este estudio. La primera contiene información de todos los estudiantes que realizaron las pruebas estandarizadas Saber11 en el año 2016 aplicadas a los estudiantes del último año escolar por el ICFES (Instituto Colombiano para el Fomento de la Educación Superior). De ella se obtiene el desempeño individual de cada estudiante para lectura y matemáticas, así como el ingreso familiar y la educación de ambos padres con los

que se estima el índice de condiciones socioeconómicas y culturales de la familia del estudiante. La segunda base de datos es el catálogo unificado de establecimientos educativos a nivel nacional que contiene información útil para la caracterización del colegio (sector, zona, etc.). El 76% de esos colegios son oficiales. Esta base es elaborada por el DANE (la autoridad estadística del gobierno nacional) a través de un censo realizado por registro administrativo. La tercera base de datos es el Índice Sintético de Calidad Educativa (ISCE) que fue descrito previamente y que solamente tiene acceso público para colegios oficiales. El ISCE se presenta para distintos niveles de educación, pero en este estudio sólo se tiene en cuenta el nivel de educación media (años décimo y onceavo). La base de datos final agrega 139,045 estudiantes matriculados en 1.320 escuelas públicas de 4 departamentos (Cundinamarca, incluida Bogotá, Antioquia, Valle del Cauca y Atlántico).

Como opción metodológica para dar cuenta de las variables de entorno se conforman 8 submuestras (4 departamentos en los que se distingue la capital del resto de municipios) y para reducir la heterogeneidad de los DMUs se aplican dos criterios de exclusión: 1. estudiantes que declararon alguna situación de discapacidad y 2. escuelas que tienen menos de 30 estudiantes matriculados presentando la prueba de Saber11. La Tabla 1 resume el número de estudiantes y escuelas por grupo de análisis.

Tabla 1. Descripción de la muestra (escuelas y estudiantes por subregión).
Colombia, 2016

Depto (grupo)	Capital		Resto		Total general	
	Escuelas	Estudiantes	Escuelas	Estudiantes	Escuelas	Estudiantes
Bogotá D.C. y Cundinamarca	320	43,109	165	16,780	485	59,889
Antioquia	163	13,644	276	25,855	439	39,499
Valle del Cauca	75	9,184	136	13,429	211	22,613
Atlántico	114	9,465	71	7,579	185	17,044
Total general	672	75,402	648	63,643	1,320	139,045

Fuente: Cálculos propios

De las 139,045 observaciones, la mayoría de los estudiantes analizados se encuentran en Bogotá D.C. y Cundinamarca (43,0% de las observaciones), seguido por Antioquia y Medellín (28,3%), Cali y Valle del Cauca (16,2%) y Barranquilla y Atlántico (12,5%). Las capitales concentran el 54.2% de los estudiantes, siendo Bogotá frente a Cundinamarca el de mayor concentración (72.0%), seguido de Barranquilla (55.5%), Cali (40.6%) y Medellín (34.5%).

La Tabla 2 presenta los descriptivos de los inputs y los outputs para las 8 submuestras. En promedio, los más altos desempeños en Lectura los alcanzan Medellín y Bogotá (capitales), y Cundinamarca y Valle (resto). En Matemáticas lo hacen Bogotá y Medellín, y Cundinamarca y Valle, respectivamente. Los inputs, por su parte, muestran las mejores condiciones socioeconómicas promedio (x_1 y x_3) en Barranquilla y Medellín (capital) y Atlántico y Valle (resto), y la más alta calidad promedio de los colegios (x_2) en Bogotá y Medellín (capital), y Atlántico y Cundinamarca (resto), respectivamente.

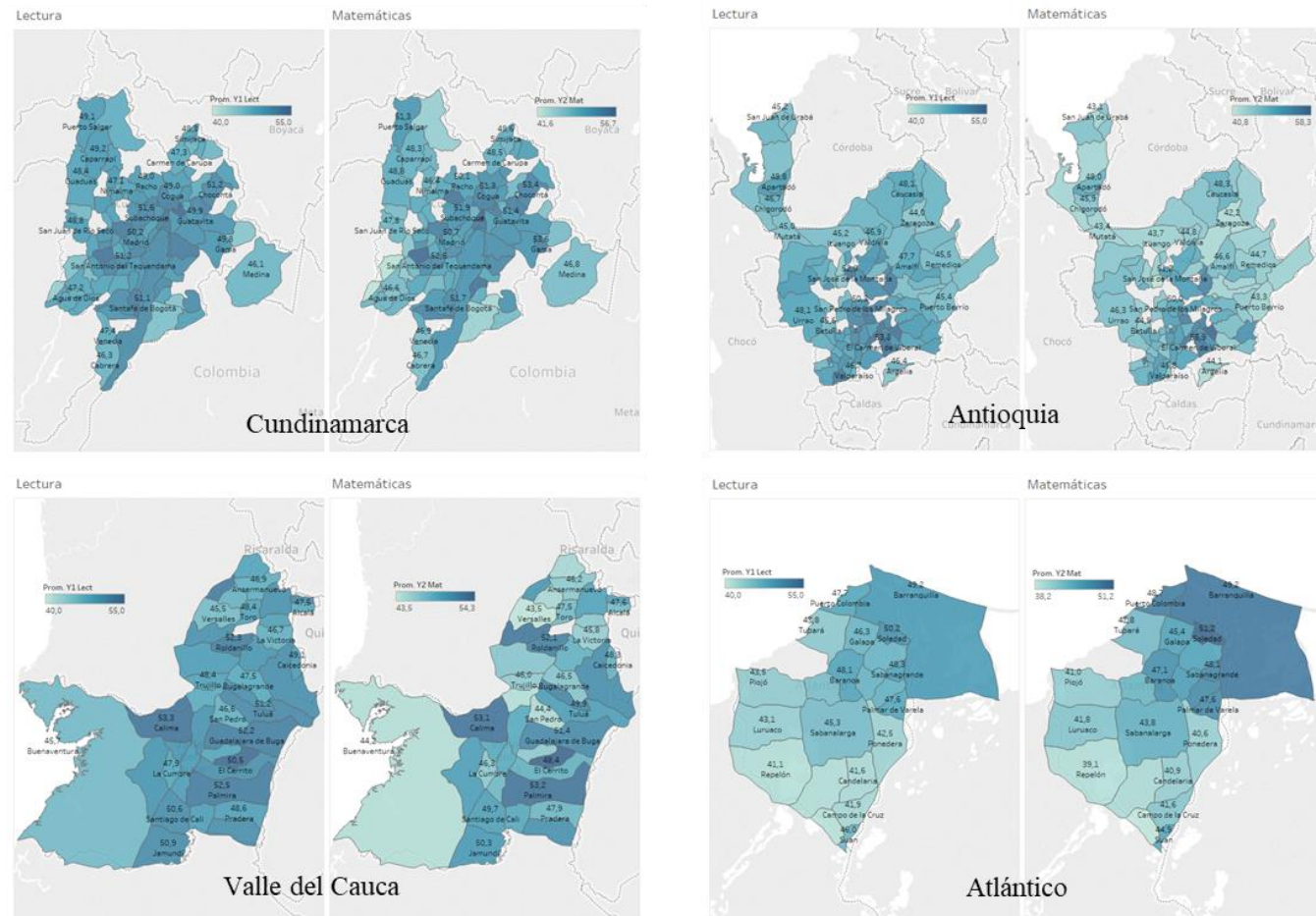
Tabla 2. Descriptivos: *inputs* y *outputs* por subregiones (Colombia, 2016)

			Ciudad Capital				Resto de las ciudades			
Bogotá D.C y Cundinamarca										
Nivel	Variable	Descripción	Min	Max	Promedio	Desvest	Min	Max	Promedio	Desvest
Estudiante	y1	Puntaje en Lectura	8.00	94.00	51.13	8.28	16.00	100.00	49.83	8.23
Estudiante	y2	Puntaje en Matemática	11.00	100.00	51.71	10.30	16.00	100.00	50.56	10.28
Estudiante	x1	Nivel socioeconómico y cultural del estudiante	4.23	6.60	5.06	0.32	4.25	6.50	4.95	0.32
Escuela	x2	Índice Sintético de Calidad	3.81	9.19	6.34	1.48	3.75	7.88	6.15	1.51
Escuela	x3	Promedio del Nivel socioeconómico y cultural del estudiante	4.57	5.67	5.06	0.14	4.666	5.676	4.952	0.127
Antioquia/Medellín										
Estudiante	y1	Puntaje en Lectura	16.00	100.00	51.29	8.69	16.00	94.00	49.39	8.85
Estudiante	y2	Puntaje en Matemática	13.00	100.00	50.57	10.52	8.00	100.00	48.93	11.18
Estudiante	x1	Nivel socioeconómico y cultural del estudiante	4.26	6.50	5.07	0.34	4.23	6.42	4.96	0.35
Escuela	x2	Índice Sintético de Calidad	3.81	7.81	6.27	1.46	3.75	8.12	6.11	1.54
Escuela	x3	Promedio del Nivel socioeconómico y cultural del estudiante	4.53	5.41	5.07	0.18	4.57	5.54	4.96	0.18
Valle del Cauca/ Cali										
Estudiante	y1	Puntaje en Lectura	4.00	94.00	50.62	8.86	9.00	100.00	49.83	8.83
Estudiante	y2	Puntaje en Matemática	7.00	100.00	49.71	10.64	13.00	100.00	49.07	10.95
Estudiante	x1	Nivel socioeconómico y cultural del estudiante	4.17	6.38	5.04	0.33	4.23	6.40	5.00	0.34
Escuela	x2	Índice Sintético de Calidad	3.81	7.68	6.25	1.50	3.70	7.83	5.39	1.48
Escuela	x3	Promedio del Nivel socioeconómico y cultural del estudiante	4.77	5.47	5.04	0.17	4.58	5.46	5.00	0.16
Atlántico/Barranquilla										
Estudiante	y1	Puntaje en Lectura	8.00	94.00	49.20	8.89	16.00	85.00	47.41	8.63
Estudiante	y2	Puntaje en Matemática	7.00	100.00	49.20	10.99	7.00	100.00	47.24	10.58
Estudiante	x1	Nivel socioeconómico y cultural del estudiante	4.21	6.42	5.13	0.36	4.24	6.45	5.05	0.33
Escuela	x2	Índice Sintético de Calidad	3.88	7.81	6.12	1.53	3.81	7.68	6.32	1.49
Escuela	x3	Promedio del Nivel socioeconómico y cultural del estudiante	4.62	5.78	5.13	0.24	4.64	5.40	5.05	0.18

Fuente: Cálculos propios

La Figura 4 presenta los mapas de densidad de los puntajes promedio en lectura y matemáticas a nivel municipal. Los municipios que no cumplen los criterios de exclusión del diseño muestral no aparecen reportados en el mapa. En ellos, se notan puntajes promedio más altos en torno a las ciudades capitales.

Figura 4. Desempeño municipal en lectura y matemáticas (promedio por área, Saber 11, ICFES, Colombia 2016).



Fuente: Elaboración propia

5. Resultados y discusión

Esta sección reporta y comenta los resultados obtenidos después de estimar una misma función de producción de la educación que maximiza el desempeño académico de los estudiantes a partir de las condiciones socioeconómicas y culturales de su propia familia y, a nivel del colegio, del promedio de este input y de la calidad de su propio colegio. Esto para 8 submuestras en Colombia durante el año 2016: 4 Departamentos (Cundinamarca, Antioquia, Valle del Cauca y Atlántico), distinguiendo sus ciudades capitales del resto de municipios. Por tratarse de estimaciones orientadas al output, las medidas son coeficientes de ineficiencia o de potencial de mejora. Es decir, un coeficiente mayor que uno se interpreta como la distancia relativa que puede mejorar un estudiante promedio frente a la frontera que conforman estudiantes con inputs similares a los suyos que obtienen los más altos desempeños en los outputs en esa vecindad. Un estudiante con un coeficiente igual a 1 es uno que está en la frontera (es perfectamente eficiente) y uno con coeficientes menores a 1 es uno super eficiente. El componente aleatorio de la estimación permite estos casos en los que ese estudiante, en promedio, supera la frontera.

La Tabla 3 presenta tres medidas de eficiencia (ineficiencia): 1. La ineficiencia total (OE), 2. La ineficiencia atribuible al estudiante y su grupo familiar (STE) y 3. La ineficiencia atribuible a la escuela (SCE). En ella, se observan también los distintos momentos de la distribución y su desviación estándar. A modo de ejemplo, el primer caso corresponde a la ciudad de Bogotá D.C. cuyos estudiantes, en promedio tienen un coeficiente de ineficiencia de 1,3486. Es decir que, dadas sus condiciones de inputs, existen estudiantes cuyo desempeño

es, en promedio, un 34,86%. En otras palabras, si a su desempeño por áreas (51,13 en Lectura y 51,71 en Matemáticas según la Tabla 2) se le añadiera su potencial de mejora, sus mejores resultados posibles podrían ser 68,95 y 69,74 respectivamente reconociendo sus niveles de inputs. En ese grupo, el alumno con mejor coeficiente OE (mínima ineficiencia igual a 0,7828) rinde 21,72% más que los comparables a él y el del peor coeficiente OE (máxima ineficiencia igual a 4,3847) casi que podría cuadruplicar su desempeño dados sus inputs.

Siguiendo con el ejemplo de Bogotá, el STE es igual a 1,1843 y el SCE a 1,1388 ($STE > SCE$). En otras palabras, en promedio, es mayor la ineficiencia atribuible a los estudiantes (18,43%) que a las escuelas (13,88%) y aunque ambos tienen un amplio potencial de mejora, se requiere un esfuerzo mayor de los estudiantes y de sus familias que del colegio.

Entre las ciudades capitales, Bogotá es la que mayor OE reporta, seguida de Medellín, Cali y Barranquilla con un rango significativo (entre 1,3486 y 1,2088). Es decir que Barranquilla, es la ciudad donde sus alumnos y sus escuelas mejor utilizan sus inputs. Entre los departamentos sin sus ciudades capitales (en adelante, los Departamentos), Antioquia es el más ineficiente, seguido de Valle y Cundinamarca y por último Atlántico, igual que su capital.

Al descomponer los efectos, la mayor ineficiencia atribuible a los estudiantes (STE) se reporta en Cali entre las ciudades y Atlántico entre los departamentos. Frente a la ineficiencia atribuible a las escuelas (SCE), las mayores ineficiencias las reportan Bogotá y Antioquia.

Tabla 3. Coeficientes de eficiencia por efectos: OE=STExSCE

Capital	Depto (grupo)	Efecto Total (OE)						
		Media geométr	Min	25%	50%	75%	Max	Desvest
Capital	Bogotá D.C. y Cundinamarca	1.3486	0.7828	1.2221	1.3462	1.4852	4.3847	0.2063
	Antioquia	1.2996	0.7530	1.1697	1.2937	1.4350	2.7999	0.2129
	Valle del Cauca	1.2538	0.7466	1.1271	1.2524	1.3891	2.8663	0.2038
	Atlántico	1.2088	0.7086	1.0876	1.2036	1.3337	3.0066	0.1926
Resto	Bogotá D.C. y Cundinamarca	1.2721	0.6924	1.1482	1.2678	1.4015	2.6913	0.1967
	Antioquia	1.3050	0.7363	1.1686	1.3014	1.4504	3.7292	0.2163
	Valle del Cauca	1.2747	0.6988	1.1465	1.2695	1.4125	2.7464	0.2108
	Atlántico	1.2140	0.6726	1.0912	1.2088	1.3430	2.6875	0.1956
		Efecto Estudiante (STE)						
		Media geométr	Min	25%	50%	75%	Max	Desvest
Capital	Bogotá D.C. y Cundinamarca	1.1843	0.7360	1.0616	1.1970	1.3026	3.7513	0.1795
	Antioquia	1.1823	0.7523	1.0523	1.1958	1.3013	2.4750	0.1861
	Valle del Cauca	1.1951	0.6816	1.0636	1.2091	1.3208	3.6838	0.1902
	Atlántico	1.1763	0.7638	1.0471	1.1892	1.2936	2.4445	0.1812
Resto	Bogotá D.C. y Cundinamarca	1.1753	0.7612	1.0482	1.1881	1.2931	2.5046	0.1800
	Antioquia	1.1775	0.6976	1.0437	1.1913	1.3003	2.7683	0.1877
	Valle del Cauca	1.1870	0.7091	1.0535	1.2012	1.3111	2.6425	0.1906
	Atlántico	1.1883	0.7464	1.0592	1.2018	1.3075	2.6875	0.1861
		Efecto Escuela (SCE)						
		Media geométr	Min	25%	50%	75%	Max	Desvest
Capital	Bogotá D.C. y Cundinamarca	1.1388	0.5006	1.0940	1.1356	1.1815	1.8684	0.0788
	Antioquia	1.0992	0.6392	1.0419	1.0930	1.1487	1.7500	0.0886
	Valle del Cauca	1.0491	0.3224	1.0049	1.0415	1.0877	1.6622	0.0741
	Atlántico	1.0276	0.5563	0.9872	1.0227	1.0662	1.7397	0.0758
Resto	Bogotá D.C. y Cundinamarca	1.0823	0.4476	1.0371	1.0777	1.1219	1.8498	0.0774
	Antioquia	1.1083	0.5821	1.0528	1.0986	1.1612	1.7880	0.0921
	Valle del Cauca	1.0738	0.6123	1.0191	1.0657	1.1255	1.8097	0.0927
	Atlántico	1.0216	0.6384	0.9779	1.0163	1.0621	1.6526	0.0836

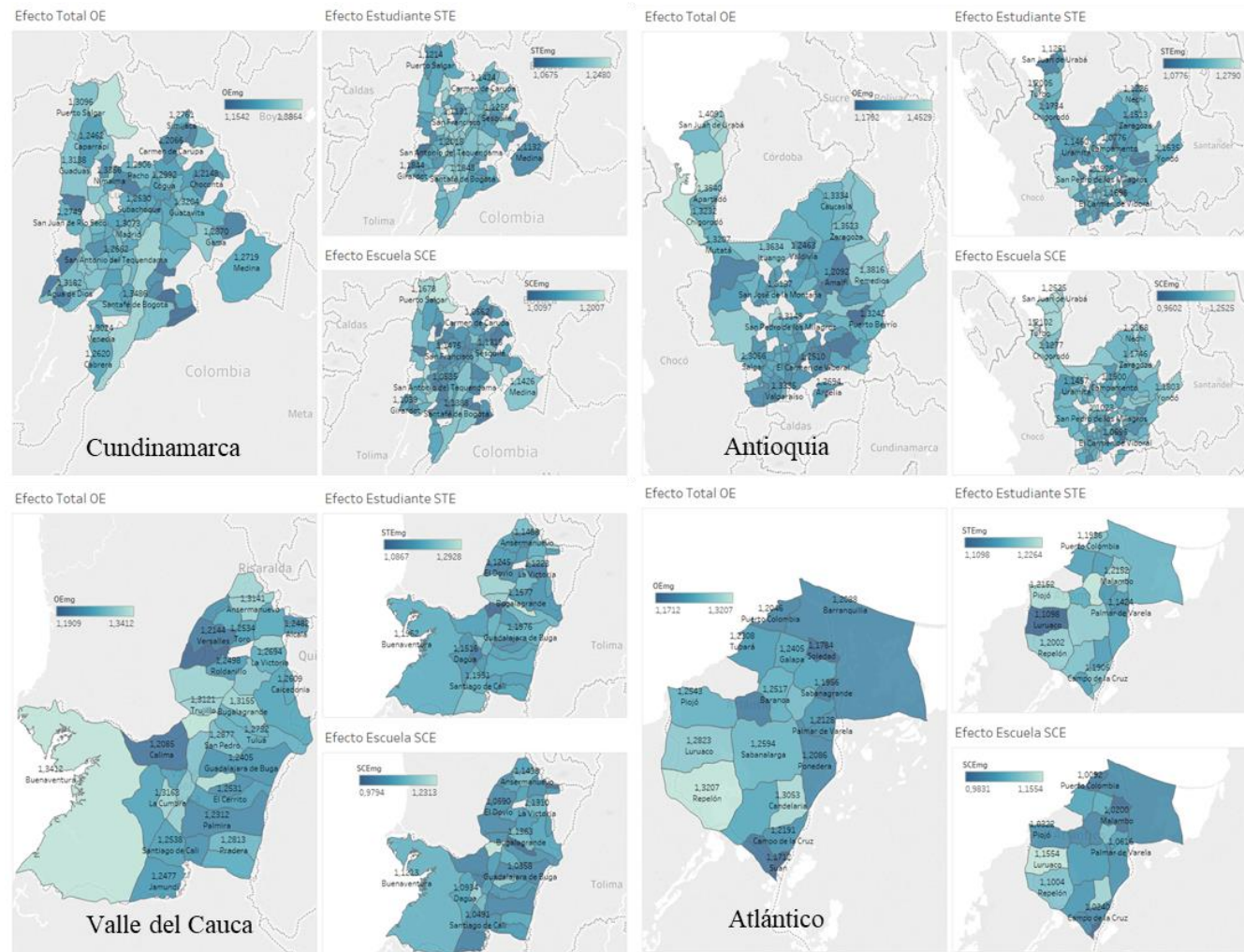
Fuente: Cálculos propios

Al revisar las distribuciones, se nota una consistencia en la mayoría de los casos. Esto es que los valores de toda la distribución se mueven uniformemente por encima o por debajo de la media geométrica de la submuestra de comparación. Por ejemplo, al estimar OE, cada

cuartil en Medellín está por debajo del correspondiente cuartil en Bogotá, al igual que sus medias geométricas. No se reportan casos inusuales.

La Figura 5 contiene los mapas de los 4 departamentos de estudio presentando, a nivel municipal, la descomposición de efectos. A diferencia de Bogotá y Cundinamarca, en las otras tres regiones la cercanía a la ciudad capital significa una ineficiencia menor, lo que se cumple aun más en el efecto escuela. La cercanía a la ciudad capital es mejor influencia para la escuela que para el alumno. Este es un hallazgo que tiene sentido en tanto los colegios están más expuestos al intercambio de docentes, prácticas pedagógicas, la presencia institucional de la ciudad capital con efectos positivos sobre su calidad, mientras sus alumnos están más expuestos a la movilidad de sus familias (y la propia) en busca de oportunidades laborales con efectos negativos. Las regiones tienen un desempeño diferente entre ellas, pero además una diferencia en su interior entre sus municipios.

Figura 5. Descomposición de efectos OE=STExSCE (Colombia 2016).



Fuente: Elaboración propia

La Tabla 4 resume los resultados anteriores y añade dos indicadores adicionales: Primero, el porcentaje de estudiantes ineficientes como un indicador más ácido de la concentración de la distribución. Consistentemente con lo anterior, el mayor porcentaje de estudiantes ineficientes se encuentra en Bogotá. Por contraste, Barranquilla y Atlántico, tienen el mayor porcentaje de estudiantes supereficientes (10,9% y 10,7%, respectivamente). Por efectos, mientras en Bogotá y Cundinamarca y Medellín y Antioquia el porcentaje de ineficientes es mayor en torno al efecto escuela, en Barranquilla y Atlántico es menor.

Segundo, esta tabla presenta el peso relativo de cada efecto. Por tratarse de coeficientes en forma de índice, además de calcular sus valores centrales con una media geométrica, sus pesos relativos se calculan como el cociente de sus logaritmos neperianos: $\ln(\text{STE})/\ln(\text{OE})$ para el efecto estudiante y $\ln(\text{SCE})/\ln(\text{OE})$ para el efecto escuela. Por ejemplo, en el primer caso, Bogotá, el peso relativo del efecto estudiante se calcula como $\ln(1,1843)/\ln(1,3486)$.

Tabla 4. Resumen descomposición de efectos OE=STExSCE

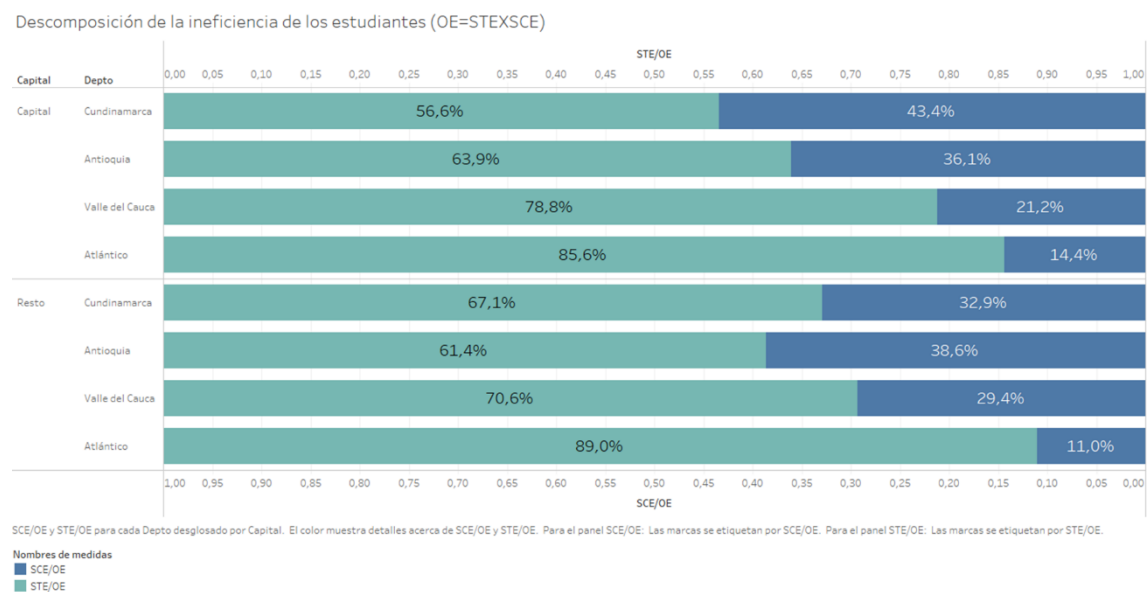
Capital	Depto (grupo)	Efecto Total		Efecto Estudiante			Efecto Escuela		
		OE	% estudiantes ineficientes	STE	% estudiantes ineficientes	Peso relativo del efecto	SCE	% estudiantes ineficientes	Peso relativo del efecto
Capital	Bogotá D.C. y Cundinamarca	1.3486	97.8%	1.1843	85.7%	56.6%	1.1388	98.0%	43.4%
	Antioquia	1.2996	95.1%	1.1823	83.7%	63.9%	1.0992	92.9%	36.1%
	Valle del Cauca	1.2538	92.0%	1.1951	85.8%	78.8%	1.0491	79.1%	21.2%
	Atlántico	1.2088	89.1%	1.1763	83.0%	85.6%	1.0276	66.4%	14.4%
Resto	Bogotá D.C. y Cundinamarca	1.2721	94.7%	1.1753	83.8%	67.1%	1.0823	91.5%	32.9%
	Antioquia	1.3050	95.4%	1.1775	82.8%	61.4%	1.1083	93.5%	38.6%
	Valle del Cauca	1.2747	93.3%	1.1870	84.2%	70.6%	1.0738	83.4%	29.4%
	Atlántico	1.2140	89.3%	1.1883	85.2%	89.0%	1.0216	61.6%	11.0%

Fuente: Cálculos propios

En todos los casos, sin excepción, el efecto estudiante domina al efecto escuela. La Figura 6 resume ambos efectos. En Barranquilla y Atlántico, además de tener las menores ineficiencias totales (OE), la composición de los efectos señala un efecto estudiante muy superior. En otras palabras, sólo el 14,4% de la ineficiencia total en Barranquilla y el 11% en el resto del Atlántico se les atribuye a las escuelas. Entre las 4 regiones, es la que más bajos puntajes tiene en los *outpts* pero al mismo tiempo es la región en la los estudiantes y, sobre todo, la escuelas, están usando mejor sus inputs y eso los coloca más cerca de su potencial.

Bogotá es el caso contrario, posee los promedios más altos en Lectura y Matemáticas a la vez que los más altos coeficientes de ineficiencia y es donde el peso relativo del efecto escuela es mayor. Antioquia y Valle, entre estos dos casos extremos, muestran que los colegios en Cali y el Valle explican en menor medida la ineficiencia total de sus alumnos.

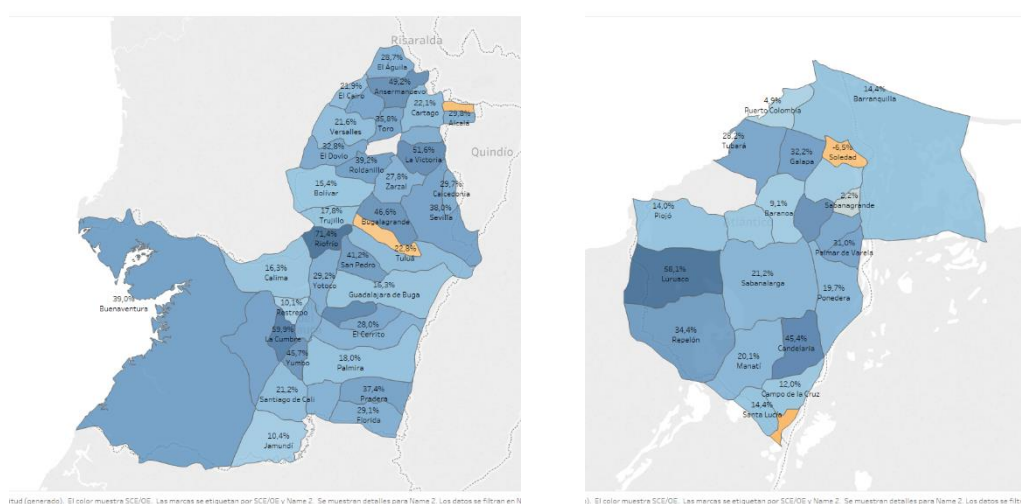
Figura 6. Peso relativo de los efectos estudiante y escuela (Colombia, 2016)



Fuente: Elaboración propia

A nivel municipal, la Figura 7 representa los casos atípicos que sólo suceden en el Valle del Cauca (Andalucía y Ulloa) y Atlántico (Soledad y Suán). En ellos, el peso relativo del SCE es negativo. Por construcción, esto sucede debido a que el SCE municipal está por debajo de 1, es decir es supereficiente, en otras palabras, utilizan eficientemente sus capacidades. En esos municipios, la totalidad de la ineficiencia se les atribuye a los estudiantes.

Figura 7. Casos atípicos en la descomposición: SCE por municipios (Valle y Atlántico, 2016)



Fuente: Elaboración propia

6. Conclusiones y Recomendaciones

Este trabajo ha tenido como objetivo estimar la eficiencia de las escuelas públicas en Colombia distinguiendo cuánto de esta eficiencia es atribuible a los estudiantes y cuánto a las escuelas. Teóricamente, es un asunto relevante porque a pesar de ser un tema con

abundantes publicaciones, los refinamientos conceptuales y de las técnicas permitan revisarlo con mayor precisión. Empíricamente, dados los requerimientos computacionales en este caso, hay pocos precedentes que utilizan estas técnicas con volúmenes de datos comparables. Desde el punto de vista de los hacedores de política permite cuestionar las decisiones de asignación de recursos a nivel sectorial y territorial.

Las tres principales conclusiones son: primero, la estimación del potencial de mejora efectivo de los estudiantes y los colegios (ineficiencia) con valores entre 1.2088 (OE Atlántico) y 1.3486 (OE Bogotá). En otras palabras, con las dotaciones actuales, el desempeño educativo, según el caso, podría ser, en promedio, al menos 20,8% superior al que actualmente consiguen los estudiantes y los colegios. Segundo, la ineficiencia del resto de los municipios es mayor al de la capital de cada región (excepto Bogotá). Esto se interpreta como un mayor potencial de mejora que podría estar asociado también a las menores capacidades institucionales de la oferta en los municipios y a la menor capacidad de referenciación y de competencia, toda vez que la oferta pública es menor en los municipios más pequeños. Tercero, tal y como lo sugiere la literatura, el efecto estudiante (STE) es estrictamente mayor al efecto escuela (SCE) en las cuatro regiones analizadas, alcanzando valores entre 56,6% y 89,0%. Esto es, sin descuidar las inversiones al nivel de la escuela, hay señales que invitan a que las intervenciones educativas incluyan con más cuidado los entornos familiares de los estudiantes.

Aunque este trabajo ha elegido tratar las variables de entorno a través de la conformación de submuestras comparables, la función de producción educativa es una básica que bien podría incluir determinantes adicionales (tanto al nivel de inputs como al de los outputs). Con ello, otras posibilidades de análisis de variables de entorno son posibles e

incluso un mejor desglose de los determinantes de cada uno de los efectos parciales. Como la literatura lo sugiere, una línea de investigación futuro con mucho potencial es la que reconoce que no todos los estudiantes aprovechan los recursos como lo hace el estudiante promedio y, por tanto, aunque los niveles de eficiencia se mejoren, es muy probable que las brechas entre estudiantes de alto y bajo desempeño se acrecienten. Por ello, la inclusión de externalidades (*bad-outputs*) es también deseable.

Referencias

- Angrist, J., Pathak, P. A., & Walters, C. R. (2013). Explaining charter school effectiveness. *American Economic Journal: Applied Economics*, 5(4), 1–27.
- Barro, R. (2001). Education and Economic Growth. In J. F. Helliwell (Ed.), *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being*. OECD.
- Bowles, S. (1970). *Towards an Educational Production Function*. (W. L. Hansen, Ed.), *Education, income, and human capital* (Vol. I). NBER National Bureau of Economic Research.
- Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106(1), 1–25.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
- Coleman, J. S. (1966). *Equality of Educational Opportunity*. Washington.
- Cordero, J. M., Prior, D., & Simancas, R. (2015). A comparison of public and private schools in Spain using robust nonparametric frontier methods. *Central European Journal of Operations Research*, 24(3), 659–680.
- Daraio, C., & Simar, L. (2005). Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach. *Journal of Productivity Analysis*, 24(1), 93–121.
- Daraio, C., & Simar, L. (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications* (Vol. 4). Boston, MA: Springer US.
- de Jorge-Moreno, J., Díaz, J., Rodríguez, D. V., & Segura, J. M. (2018). Analysis of educational efficiency and its explanatory factors considering the effect of ownership in Colombia with Pisa 2012 data. *Revista Desarrollo Sociedad*, 80, 89–118.
- De Witte, K., & López-Torres, L. (2015). Efficiency in Education . A review of literature and a way forward. *Journal of the Operational Research Society*, 68, 1–25.
- Deprins, D., Simar, L., & Tulkens, H. (1984). Measuring labor inefficiency in post offices. In H. Marchand, M., Pestieau, P., Tulkens (Ed.), *The Performance of Public Enterprises: Concepts and Measurements* (pp. 243–267). North Holland.
- Gregorio, J. De, & Lee, J.-W. (2002). Education and Income Inequality: New Evidence From Cross-Country Data. *Review of Income and Wealth*, 48(3), 395–416.
- Hanushek, E. A. (1979). Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *The Journal of Human Resources*, 14(3), 351–388.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3), 1141–1177.

- Hanushek, E. A. (2006). School Resources. In Eric A. Hanushek & Finis Welch (Eds.), *Handbook of the Economics of Education* (Vol. 2, pp. 865–908).
- Hanushek, E. A. (2008). Education Production Functions. In *The New Palgrave Dictionary of Economics* (pp. 1–5). London: Palgrave Macmillan UK.
- Hanushek, E. A., & Kimko, D. D. (2000). Schooling, Labor Force Quality, and the Growth of Nations. *American Economic Review*, 90(5), 1184–1208.
- Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3), 607–668.
- Hanushek, E. A., & Wößmann, L. (2010). Education and Economic Growth Early Studies of Schooling Quantity and Economic Growth. In D. J. Brewer & P. J. McEwan (Eds.), *Economics of Education* (pp. 60–67). Elsevier Ltd.
- Martins, L., & Veiga, P. (2010). Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review*, 29(6), 1016–1033.
- Muhammad, R., Iqbal, N., & Tasneem, S. (2015). The influence of Parents Educational level on Secondary School Students Academic achievements in District Rajanpur. *Journal of Education and Practice*, 6(16), 76–79.
- OECD. (2017). *Education at a Glance 2017* (Education at a Glance). OECD Publishing. <https://doi.org/10.1787/eag-2017-en>
- Psacharopoulos, G. (1981). Returns to Education: An Updated International Comparison. *Comparative Education*, 17(3), 321–341.
- Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics*, 12(2), 111–134.
- Silva, M. C. A., & Thanassoulis, E. (2001). Decomposing school and school-type efficiency. *European Journal of Operational Research*, 132(2), 357–373.
- Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3), 391–424.
- Tauchmann, H. (2012). Partial frontier efficiency analysis. *Stata Journal*, 12(3), 461–478.
- Tenenhaus, M., & Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, 50(1), 91–119.
- Thanassoulis, E., & Silva, M. C. A. (2002). School outcomes: sharing the responsibility between pupil and school. *Education Economics*, 10(2), 183–207.
- Thieme, C., Prior, D., & Tortosa-Ausina, E. (2013). A multilevel decomposition of school performance using robust nonparametric frontier techniques. *Economics of Education Review*, 32(1), 104–121.
- Thieme, C., Prior, D., Tortosa-Ausina, E., & Gempp, R. (2016). Value added, educational accountability approaches and their effects on schools' rankings: Evidence from Chile. *European Journal of Operational Research*, 253(2), 456–471.